

# Estimation for Dynamic and Static Panel Probit Models with Large Individual Effects

Wei Gao<sup>\*</sup>      Wicher Bergsma<sup>\*</sup>      Qiwei Yao<sup>\*,†</sup>

<sup>\*</sup>Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics,  
Northeast Normal University, Changchun, Jilin 130024, China

<sup>\*</sup>Department of Statistics, London School of Economics, London, UK

<sup>†</sup>Guanghua School of Management, Peking University, Beijing, China

## Abstract

For discrete panel data, the dynamic relationship between successive observations is often of interest. We consider a dynamic probit model for short panel data. A problem with estimating the dynamic parameter of interest is that the model contains a large number of nuisance parameters, one for each individual. Heckman proposed to use maximum likelihood estimation of the dynamic parameter, which, however, does not perform well if the individual effects are large. We suggest new estimators for the dynamic parameter, based on the assumption that the individual parameters are random and possibly large. Theoretical properties of our estimators are derived and a simulation study shows they have some advantages compared to Heckman's estimator.

**Key Words:** Dynamic probit regression; Generalized linear models; Panel data; Probit models; Static probit regression.

## 1 Introduction

Short binary-valued time series in the presence of covariates are often available in panel studies for which observations are taken on a panel of individuals over a short time period. Dynamic probit regression is one of the most frequently used statistical models to analyse this type of data. To set the scene, consider a panel of  $n$  independently sampled individuals. For each individual  $i$ , binary observations, denoted by  $d_{i1}, \dots, d_{iT}$ , are taken at time  $1, \dots, T$ , and the observations are assumed

to satisfy the latent dynamic model:

$$d_{i1} = I(\tau_i + \mathbf{x}_{i1}'\boldsymbol{\beta} + \epsilon_{i1} > 0), \quad d_{it} = I(\tau_i + \gamma d_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \epsilon_{it} > 0) \quad \text{for } 1 < t \leq T, \quad (1)$$

where  $I(\cdot)$  denotes the indicator function,  $\{\epsilon_{it}\}$  are independently and identically distributed with mean 0 and variance 1,  $\{\mathbf{x}_{it}\}$  are  $k \times 1$  covariate vectors,  $\tau_i$  is an unknown intercept representing the  $i$ -th individual effect, and the autoregressive coefficient  $\gamma$  and the regressive coefficient  $\boldsymbol{\beta}$  are unknown parameters which are assumed to be the same for all individuals. In (1), only the  $d_{it}$  and  $\mathbf{x}_{it}$  are observable. The goal is often to estimate  $\gamma$  and  $\boldsymbol{\beta}$  while the  $\tau_i$  are treated as nuisance parameters. As with most panel data, the number of individuals  $n$  is large while the length of observed time period  $T$  is small. Therefore the asymptotic approximations are often derived with  $n \rightarrow \infty$  and  $T$  fixed.

Model (1) is a dynamic panel probit regression model, as the dynamic dependence is reflected by the autoregressive parameter  $\gamma$  which links  $d_{it}$ , i.e. the state at time  $t$ , to the state at time  $t - 1$ . When  $\gamma = 0$ , (1) reduces to a static panel probit regression, as now  $d_{it}$  is independent of  $d_{i,t-1}, d_{i,t-2}, \dots$ . Model (1) has been used for various applications in microeconomics by, among others, Heckman (1978), Arellano and Honore (2001), and Hsiao (2003, Section 7.5). For example, Heckman (1978, 1980) used model (1) to reveal some interesting dynamics in unemployment data:  $d_{it} = 0$  indicates that individual  $i$  is unemployed at time  $t$ , and 1 otherwise, while the covariate  $\mathbf{x}_{it}$  stands for the factors (such as age, education, family background etc) which may affect the employment status. These studies tried to provide statistical evidence to answer questions such as: *Does current unemployment cause future unemployment?* If  $\gamma > 0$  this indicates that being in employment at time  $t$  increases the chances of being in employment at time  $t + 1$ .

Various estimation methods have been proposed for model (1). By treating the individual effects  $\tau_1, \dots, \tau_n$  as nuisance parameters or incidental parameters (Neyman and Scott, 1948), Heckman (1980) adopted the maximum likelihood estimator of  $\gamma$  as well as  $\boldsymbol{\beta}$  when  $\epsilon_{it}$  are normally distributed. Chamberlain (1980, 1985), Honore and Kyriazidou (2000), and Lancaster (2002) considered the models with logistic distributed  $\epsilon_{it}$ . They proposed a consistent estimator of  $\gamma$  and derived its convergence rate.

Bartolucci and Farcomeni (2009) and Bartolucci and Nigro (2010) considered some extended versions of dynamic logit models with heterogeneity beyond those reflected by the covariates in the models. A standard method to deal with incidental parameter problems is to use a conditional likelihood to eliminate the incidental parameters by conditioning on sufficient statistics for those parameters; see, e.g. Chamberlain (1980), Bartolucci and Nigro (2010), and also Lancaster (2000).

An attractive alternative is to treat individual effects  $\tau_i$  as random effects with pre-specified priors. But as far as we are aware, most literature on panel probit regression taking this approach only deal with the static model (i.e.  $\gamma = 0$  in (1)) only. For example, Chamberlain (1980, 1985) discussed the maximum likelihood estimator for  $\beta$  with a given prior distribution for  $\tau_i$ . Arellano and Bonhomme (2009) showed that this estimator is robust with respect to the choice of prior when  $T$  is large. Manski (1987) proposes maximum score methods to estimate  $\beta$  when the distribution of the errors is unknown and  $\gamma$  is equal to zero for model (1). Smoothed maximum score estimators were developed by Horowitz (1992). See also Arellano (2003) for a survey of static probit models.

In this paper, we propose new estimation methods for  $\gamma$  and  $\beta$  in model (1) with  $\epsilon_{it} \sim N(0, 1)$ . We treat  $\tau_i$  as random effects but with an unspecified prior. Our methods are designed for the cases when the individual effects  $\tau_1, \dots, \tau_n$  are large while  $T$  is small. Note that when  $\tau_i$  are large, there is an innate difficulty in estimating  $\gamma$  and  $\beta$  as the outcome of the random event  $\{\tau_i + \gamma d_{i,t-1} + \mathbf{x}'_{it}\beta + \epsilon_{it} > 0\}$  may be dominated by the value of  $\tau_i$ . In fact Heckman (1980) reported that the maximum likelihood estimator for  $\gamma$  behaved poorly when the variance of  $\tau_i$  is large; see Table 4.2 in Heckman (1980). Furthermore, our simulation results indicate that our methods work as well as Heckman's (1980) method when the variance of  $\tau_i$  are, for example, equal to 1 and 4.

The rest of the paper is organised as follows: Section 2 presents the new estimation methods together with their asymptotic properties. For the simplicity of the presentation, we consider the case  $T = 2$  only, though the methods can be extended to the cases with  $T > 2$ . Simulations are reported in Section 3 and an example is analyzed

in Section 4. Some technical proofs are relegated to the Appendix.

## 2 Estimation methods

We consider model (1) with  $T = 2$ , namely

$$d_{i1} = I(\tau_i + \mathbf{x}'_{i1}\boldsymbol{\beta} + \epsilon_{i1} > 0), \quad d_{i2} = I(\tau_i + \gamma d_{i1} + \mathbf{x}'_{i2}\boldsymbol{\beta} + \epsilon_{i2} > 0), \quad i = 1, \dots, n, \quad (2)$$

where  $\{\epsilon_{i1}\}$  and  $\{\epsilon_{i2}\}$  are independent and  $N(0, 1)$ , and  $\tau_i$  is independent of  $\epsilon_{i1}$  and  $\epsilon_{i2}$ . Furthermore, we assume that  $\{\tau_i\}$  are independent with a common density function  $f(\cdot)$  which satisfies condition C1 below.

**C1** The density function of  $\tau_i$  admits the expression

$$f(x) = \frac{1}{\sigma_\tau} h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right), \quad (3)$$

where  $h(\cdot)$  is a density function with mean 0 and variance 1,  $h(x)$  is continuous at  $x = 0$ , and  $\mu_\tau$  and  $\sigma_\tau > 0$  are constants.

We present below the new estimation methods for the three scenarios: (i) estimating the autoregressive coefficient  $\gamma$  without covariates (i.e.  $\boldsymbol{\beta} = 0$ ), (ii) estimating regressive coefficient vector  $\boldsymbol{\beta}$  for the static model (i.e.  $\gamma = 0$ ), and (iii) estimating  $\gamma$  and  $\boldsymbol{\beta}$  together. All those methods are derived based on some asymptotic arguments when  $\sigma_\tau \rightarrow \infty$ , and therefore the methods are particularly relevant when the individual effects are large.

### 2.1 Estimation of $\gamma$ when $\boldsymbol{\beta} = 0$

When  $\boldsymbol{\beta} = 0$ , model (2) reduces to

$$d_{i1} = I(\tau_i + \epsilon_{i1} > 0), \quad d_{i2} = I(\tau_i + \gamma d_{i1} + \epsilon_{i2} > 0), \quad i = 1, \dots, n. \quad (4)$$

As  $\tau_i$ ,  $\epsilon_{i1}$  and  $\epsilon_{i2}$  are independent, and  $\epsilon_{i1}$  and  $\epsilon_{i2}$  are  $N(0, 1)$ , it holds that

$$\begin{aligned} P\{d_{i1} = 0, d_{i2} = 0\} &= \int \Phi(-x)\Phi(-x)f(x)dx, \\ P\{d_{i1} = 0, d_{i2} = 1\} &= \int \Phi(-x)\Phi(x)f(x)dx, \\ P\{d_{i1} = 1, d_{i2} = 0\} &= \int \Phi(x)\Phi(-x - \gamma)f(x)dx, \\ P\{d_{i1} = 1, d_{i2} = 1\} &= \int \Phi(x)\Phi(x + \gamma)f(x)dx, \end{aligned} \quad (5)$$

where  $\Phi$  is the standard normal distribution function, and  $f(\cdot)$  is the density function of  $\tau_i$ . We state in Proposition 1 below an asymptotic property on the ratio of the two probabilities listed above, on which our new estimation method for  $\gamma$  is motivated. Its proof follows immediately from Lemmas 1 and 2 in the Appendix.

**Proposition 1.** Under condition C1, it holds that

$$\begin{aligned} \lim_{\sigma_\tau \rightarrow \infty} \frac{P\{d_{i1} = 1, d_{i2} = 0\}}{P\{d_{i1} = 0, d_{i2} = 1\}} &= \lim_{\sigma_\tau \rightarrow \infty} \frac{\int \Phi(x)\Phi(-x - \gamma)f(x)dx}{\int \Phi(-x)\Phi(x)f(x)dx} \\ &= \frac{\int \Phi(x)\Phi(-x - \gamma)dx}{\int \Phi(x)\Phi(-x)dx} = G(\gamma), \end{aligned} \quad (6)$$

where

$$G(\gamma) = -\sqrt{\pi}\gamma\Phi\left(-\frac{\gamma}{\sqrt{2}}\right) + \exp\left\{-\frac{\gamma^2}{4}\right\}. \quad (7)$$

Proposition 1 above suggests the following estimator for  $\gamma$ :

$$\hat{\gamma} = G^{-1}(\widehat{W}), \quad (8)$$

where  $G(\cdot)$  is given in (7), and

$$\widehat{W} = \sum_{i=1}^n I(d_{i1} = 1, d_{i2} = 0) / \sum_{i=1}^n I(d_{i1} = 0, d_{i2} = 1), \quad (9)$$

i.e.  $\widehat{W}$  is a plug-in estimator for the ratio of the two probabilities on the left hand side of (6). The asymptotic properties of  $\hat{\gamma}$  are stated in the theorem below. Put

$$\kappa_n = \left\{ \sum_{i=1}^n I(d_{i1} = 0, d_{i2} = 1) \right\}^{1/2}, \quad \sigma^2 = \frac{G(\gamma) + G^2(\gamma)}{[G'(\gamma)]^2} = \frac{G(\gamma) + G^2(\gamma)}{\pi\Phi^2(-\gamma/\sqrt{2})}. \quad (10)$$

**Theorem 1.** Under condition C1, the following assertions holds.

- (i)  $\lim_{\sigma_\tau \rightarrow \infty} \lim_{n \rightarrow \infty} P\{|\hat{\gamma} - \gamma| \geq \eta\} = 0$  for any  $\eta > 0$ .
- (ii)  $\lim_{n \rightarrow \infty} P\{\kappa_n(\hat{\gamma} - \gamma) \leq x\} = \Phi(x/\sigma)$  for any real number  $x$ , provided that the first derivative of  $h(\cdot)$  is continuous and  $\sigma_\tau = a\sqrt{n}$  for some constant  $a$ .

**Remark 1.** (i) Theorem 1(i) can be viewed as a version of consistency for  $\hat{\gamma}$ . Theorem 1(ii) indicates that  $\hat{\gamma}$  is asymptotically normal if we restrict  $\sigma_\tau = a\sqrt{n}$ . Note that the convergence rate  $\kappa_n$  defined in (10) admits the asymptotic relation: the standard  $\sqrt{n}$ , as

$$\kappa_n^2 = nf(\mu_\tau) \int \Phi(u)\Phi(-u)du + o_p(n/\sigma_\tau).$$

See the proof of Theorem 1 in the Appendix. Thus the larger the sample size  $n$  is, the faster  $\hat{\gamma}$  converges to  $\gamma$ .

(ii) Only two out of the four probabilities in (5) are used in defining the estimator  $\hat{\gamma}$ . Indeed the observations with  $(d_{i1}, d_{i2}) = (0, 0)$  or  $(1, 1)$  are not utilized in (8). In fact when  $\sigma_\tau$  is large, those data provide little information on  $\gamma$ , as

$$\begin{aligned} \lim_{\sigma_\tau \rightarrow \infty} P\{d_{i1} = 0, d_{i2} = 0\} &= \lim_{\sigma_\tau \rightarrow \infty} \int \Phi(-x)\Phi(-x) \frac{1}{\sigma_\tau} h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right) dx \\ &= \lim_{\sigma_\tau \rightarrow \infty} \int \Phi(-\sigma_\tau t - \mu_\tau)\Phi(-\sigma_\tau t - \mu_\tau) h(t) dt = H(0), \end{aligned}$$

and similarly

$$\lim_{\sigma_\tau \rightarrow \infty} P\{d_{i1} = 1, d_{i2} = 1\} = 1 - H(0).$$

where  $H(x)$  is cumulative distribution function of  $h(x)$ .

## 2.2 Estimation of $\beta$ when $\gamma = 0$

Let

$$D_n = \{(d_{i1}, d_{i2})' : d_{i1} + d_{i2} = 1 \text{ for } i = 1, \dots, n\}$$

and denote the number of elements in  $D_n$  by  $m$ . Without loss of generality, suppose that  $d_{i1} + d_{i2} = 1$  for  $i = 1, \dots, m$ .

We find the conditional probability

$$\begin{aligned} & P\{d_{i1} = 1, d_{i2} = 0 | d_{i1} + d_{i2} = 1, \mathbf{x}_{i1}, \mathbf{x}_{i2}\} \\ &= \frac{\int \Phi(\mathbf{x}'_{i1}\boldsymbol{\beta} + t)\Phi(-\mathbf{x}'_{i2}\boldsymbol{\beta} - t)f(t)dt}{\int \Phi(\mathbf{x}'_{i1}\boldsymbol{\beta} + t)\Phi(-\mathbf{x}'_{i2}\boldsymbol{\beta} - t)f(t)dt + \int \Phi(-\mathbf{x}'_{i1}\boldsymbol{\beta} - t)\Phi(\mathbf{x}'_{i2}\boldsymbol{\beta} + t)f(t)dt}. \end{aligned}$$

Under (3), we can similarly prove

$$\lim_{\sigma_\tau \rightarrow \infty} P\{d_{i1} = 1, d_{i2} = 0 | d_{i1} + d_{i2} = 1, \mathbf{x}_{i1}, \mathbf{x}_{i2}\} = \frac{G((\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta})}{G((\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta}) + G(-(\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta})}.$$

For sufficiently large  $\sigma_\tau$ , we can replace the conditional likelihood of  $\boldsymbol{\beta}$  given  $D_n$  by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^m p_i^{z_i} (1 - p_i)^{1-z_i} \quad (11)$$

where  $z_i = I(d_{i1} = 1, d_{i2} = 0)$  and  $1 - z_i = I(d_{i1} = 0, d_{i2} = 1)$ , and

$$p_i = \frac{G((\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta})}{G((\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta}) + G(-(\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta})}. \quad (12)$$

Note that  $p_i = K((\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta})$  for the monotone function  $K$  defined as

$$K(t) = \frac{G(t)}{G(t) + G(-t)},$$

Hence, (12) is a generalized linear model of the form

$$K^{-1}(p_i) = (\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta}.$$

So iterative reweighted least squares methods for generalized Models given by McCullagh and Nelder (1989) can be applied to (11) to estimate the parameter  $\boldsymbol{\beta}$ . Under some regularity conditions and  $\sigma_\tau \rightarrow \infty$ , consistency of  $\boldsymbol{\beta}$  can be shown.

### 2.3 Simultaneous estimation of $\gamma$ and $\boldsymbol{\beta}$

As in Section 2.2, we have

$$\lim_{\sigma_\tau \rightarrow \infty} P\{d_{i1} = 1, d_{i2} = 0 | d_{i1} + d_{i2} = 1, \mathbf{x}_{i1}, \mathbf{x}_{i2}\} = \frac{G(\gamma + (\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta})}{G(\gamma + (\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta}) + G(-(\mathbf{x}_{i2} - \mathbf{x}_{i1})'\boldsymbol{\beta})}.$$

For large  $\sigma_\tau$ , we replace the conditional likelihood of  $\gamma$  and  $\beta$  given  $D_n$  by

$$L(\beta) = \prod_{i=1}^m p_i^{z_i} (1 - p_i)^{1-z_i} \quad (13)$$

where  $z_i = I(\{d_{i1} = 1, d_{i2} = 0\})$  and  $1 - z_i = I(d_{i1} = 0, d_{i2} = 1)$ , and

$$p_i = \frac{G(\gamma + (\mathbf{x}_{i2} - \mathbf{x}_{i1})' \beta)}{G(\gamma + (\mathbf{x}_{i2} - \mathbf{x}_{i1})' \beta) + G(-(\mathbf{x}_{i2} - \mathbf{x}_{i1})' \beta)}. \quad (14)$$

Let

$$\mathbf{X}^* = (\mathbf{x}_{12} - \mathbf{x}_{11}, \mathbf{x}_{22} - \mathbf{x}_{21}, \dots, \mathbf{x}_{m2} - \mathbf{x}_{m1})$$

**Theorem 2.** (14) is identifiable for  $\gamma$  and  $\beta$  if the rank of  $\mathbf{X}^*$  is equal to  $k$  (the dimension of  $\mathbf{x}_{2i} - \mathbf{x}_{1i}$ ) and at least there exist  $j$  and  $1 \leq s_1, \dots, s_k \leq m$  such that

$$\mathbf{x}_{j2} - \mathbf{x}_{j1} = a_1(\mathbf{x}_{s_12} - \mathbf{x}_{s_11}) + a_2(\mathbf{x}_{s_22} - \mathbf{x}_{s_21}) + \dots + a_k(\mathbf{x}_{s_k2} - \mathbf{x}_{s_k1})$$

where  $a_1, \dots, a_k$  are non-positive real numbers.

The conditions in Theorem 2 are sufficient and can be satisfied with probability close to 1 for a large sample size  $n$  if the covariate  $\mathbf{x}_{i2} - \mathbf{x}_{i1}$  is a continuous variable and its covariance matrix is positive definite.

**Corollary.** Under the condition in Theorem 2, and with  $\mathbf{1}_m$  be the  $m$ -dimensional vector with all components 1, the rank of  $(\mathbf{1}_m, \mathbf{X}^{*'})$  is  $k + 1$ .

From the Corollary, it seems that the identifiability condition relating to (14) is stronger than that of linear models since that the rank of design matrix being equal to the number of parameters is sufficient for linear models to be identified.

### 3 Simulation study

In this section, we use simulations to estimate the root mean squared errors (RMSEs) of the estimators proposed in Section 2. In Table 1, RMSEs of  $\gamma$  in Model (4) are given for different distributions of the individual effects. In Table 2, RMSEs of  $\gamma$  and  $\beta$  in Model (1) are given, with the  $x_{i1}$  sampled from the standard normal distribution



Table 1: Simulated RMSEs of the new estimator of the dynamic parameter  $\gamma$  in Model (4) (100 replications)

		$n = 1000$	$n = 5000$	
Distribution of the $\tau_i$	$\gamma$	RMSE	Distribution of the $\tau_i$	RMSE
U(-3,3)	-2	0.16	U(-10,10)	0.21
	-1.5	0.24		0.19
	-1	0.23		0.14
	-0.5	0.20		0.15
	0	0.15		0.13
	0.5	0.21		0.31
	1	0.18		0.14
	1.5	0.15		0.15
	2	0.25		0.18
N(0,4)	-2	0.30	N(0,25)	0.16
	-1.5	0.15		0.19
	-1	0.20		0.13
	-0.5	0.15		0.12
	0	0.15		0.11
	0.5	0.16		0.10
	1	0.17		0.11
	1.5	0.18		0.12
	2	0.23		0.17

and  $x_{i2} = x_{i1} + N(0, 1)$ ; the individual effects are normally distributed with mean 0 and variance 2. For normally distributed individual effects with mean 0 and variance  $\sigma^2$  in Model (1), Heckman (1980) has proposed the maximum likelihood estimation of the dynamic parameter  $\gamma$  and  $\sigma^2$ . In Tables 3 and 4 the RMSE of our new estimator is compared with the RMSE of Heckman's estimator, in the former table for normally distributed individual effects and in the latter for individual effects with a mixture normal distribution. We see that our estimator is comparable to Heckman's for normally distributed effects with moderate variance, but greatly outperforms it when individual effects are mixed normal distributions.

Table 2: Simulated RMSE of new estimators of  $\gamma$  and  $\beta$  for Model (1) (200 replicates and  $n = 1000$ )

$\gamma$	$\beta$	RMSE( $\hat{\gamma}$ )	RMSE( $\hat{\beta}$ )	$\gamma$	$\beta$	RMSE( $\hat{\gamma}$ )	RMSE( $\hat{\beta}$ )
-1	0	0.20	0.08	0	-1	0.20	0.15
-0.5	0	0.17	0.08	0	-0.5	0.18	0.10
0	0	0.14	0.08				
0.5	0	0.16	0.08	0	0.5	0.16	0.10
1	0	0.16	0.09	0	1	0.19	0.13
-1	1	0.22	0.13	1	1	0.25	0.16
-0.5	0.5	0.19	0.10	0.5	0.5	0.15	0.09
0.5	-0.5	0.16	0.10	-0.5	-0.5	0.17	0.10
1	-1	0.22	0.18	-1	-1	0.24	0.13

Table 3: Comparison of RMSE of new estimator ( $\hat{\gamma}_G$ ) and Heckman's estimator ( $\hat{\gamma}_H$ ) for normally distributed individual effects (200 replicates for sample size  $n = 1000$ ).

Distribution of the $\tau_i$	$\gamma$	RMSE( $\hat{\gamma}_G$ )	RMSE( $\hat{\gamma}_H$ )	RMSE( $\hat{\sigma}_H$ )
$N(0, 1)$	-1	0.16	0.13	0.13
	-0.5	0.14	0.11	0.12
	0	0.12	0.09	0.11
	0.5	0.13	0.10	0.11
	1	0.13	0.10	0.12
$N(0, 4)$	-1	0.20	0.16	0.25
	-0.5	0.18	0.15	0.21
	0	0.15	0.12	0.18
	0.5	0.17	0.14	0.26
	1	0.17	0.15	0.20

Table 4: Comparison of new estimator ( $\hat{\gamma}_G$ ) with Heckman's ( $\hat{\gamma}_H$ ) when individual effects are distributed as  $0.5N(-6, 9) + 0.5N(6, 9)$  (200 replicates with sample size  $n = 3000$ ).

$\gamma$	RMSE( $\hat{\gamma}_G$ )	RMSE( $\hat{\gamma}_H$ )	RMSE( $\hat{\sigma}_H$ )
-1	0.37	0.81	3.81
-0.5	0.29	0.75	3.82
0	0.30	0.64	3.86
0.5	0.29	0.59	3.81
1	0.30	0.53	3.85

## 4 An example

We analyze the data set listed in Table (5) which has previously been considered by Heckman(1981). The dynamics of female labor supply is investigated based on panel data from the years 1968 to 1970, and 1971 to 1973. Model (1) is applied to estimate the dynamic parameter with  $T = 3$  and  $x_{it} \equiv 0$ . Let  $n_{ijl}$  be the number of observations of runs pattern  $(i, j, l)$  in Table (5) for  $i, j, l = 0, 1$ . As in Section 2.2, the following methods can be developed to estimate  $\gamma$ . The estimates are given in Table (6), where  $\hat{\gamma}_G$  is the new estimator and  $\hat{\gamma}_H$  and  $\hat{\sigma}_H$  are Heckman's estimators.

$$\hat{\gamma}_G = \arg \max \{p_{001}^{n_{001}} p_{010}^{n_{010}} p_{100}^{n_{100}} p_{110}^{n_{110}} p_{011}^{n_{011}} p_{101}^{n_{101}}\}$$

where

$$p_{001} = \frac{\int \Phi(-t)\Phi(-t)\Phi(t)dt}{K_1}, p_{010} = \frac{\int \Phi(-t)\Phi(t)\Phi(-t-\gamma)dt}{K_1}, p_{100} = \frac{\int \Phi(t)\Phi(-t-\gamma)\Phi(-t)dt}{K_1},$$

$$p_{110} = \frac{\int \Phi(t)\Phi(t+\gamma)\Phi(-t-\gamma)dt}{K_2}, p_{011} = \frac{\int \Phi(-t)\Phi(t)\Phi(t+\gamma)dt}{K_2}, p_{101} = \frac{\int \Phi(t)\Phi(-t-\gamma)\Phi(t)dt}{K_2},$$

and

$$K_1 = \int \Phi(-t)\Phi(-t)\Phi(t)dt + \int \Phi(t)\Phi(-t-\gamma)\Phi(-t)dt + \int \Phi(t)\Phi(-t-\gamma)\Phi(-t)dt,$$

$$K_2 = \int \Phi(t)\Phi(t+\gamma)\Phi(-t-\gamma)dt + \int \Phi(-t)\Phi(t)\Phi(t+\gamma)dt + \int \Phi(t)\Phi(-t-\gamma)\Phi(t)dt.$$

Table 5: Runs patterns in the data (1 corresponds to work in the year, 0 corresponds to no work)

Runs patterns			No. of	Runs pattern			No.of
1968	1969	1970	observations	1971	1972	1973	observations
women aged 45-59 in 1968							
0	0	0	87	0	0	0	96
0	0	1	5	0	0	1	5
0	1	0	5	0	1	0	4
1	0	0	4	1	0	0	8
1	1	0	8	1	1	0	5
0	1	1	10	0	1	1	2
1	0	1	1	1	0	1	2
1	1	1	78	1	1	1	76
women aged 30-44 in 1968							
0	0	0	126	0	0	0	133
0	0	1	16	0	0	1	13
0	1	0	4	0	1	0	5
1	0	0	12	1	0	0	16
1	1	0	24	1	1	0	8
0	1	1	20	0	1	1	19
1	0	1	5	1	0	1	8
1	1	1	125	1	1	1	130

From the analyzed results in the age group 49-59 and runs pattern from 1971 to 1973, neither Heckman's method nor the proposed method yield evidence of a dynamic relationship, and perhaps more data needs to be collected. However, the difference for the older group between the period 1968-170 and 1971-1973 is significant; the difference for the younger group between the period 1968-170 and 1971-1973 is not significant. For age group 30-44, both the proposed method and Heckman's method yield a significant dynamic relationship, with a positive estimated value of  $\gamma$  (here, positivity of  $\gamma$  implies the unsurprising result that currently holding a job increases the likelihood of holding a job in future).

Table 6: Comparison of new estimator ( $\hat{\gamma}_G$ ) with Heckman's ( $\hat{\gamma}_H$ ) for data in Table 5

panel data(1969-1970)			panel data(1971-1973)		
$\hat{\gamma}_G$ (s.e.)	$\hat{\gamma}_H$ (s.e.)	$\hat{\sigma}_H$ (s.e.)	$\hat{\gamma}_G$ (s.e.)	$\hat{\gamma}_H$ (s.e.)	$\hat{\sigma}_H$ (s.e.)
women aged 45-59 in 1968					
0.62 (0.20)	0.54 (0.27)	3.24 (0.65)	-0.16 (0.26)	-0.28 (0.36)	5.59 (1.33)
women aged 30-44 in 1968					
0.48 (0.13)	0.47 (0.17)	2.15 (0.28)	0.51 (0.14)	0.43 (0.19)	2.63 (0.37)

## Appendix

**Lemma 1.** If  $f(x)$  satisfies the conditions given in Theorem 1, then

$$\int \Phi(x)\Phi(-x-\gamma)f(x)dx = f(\mu_\tau) \int \Phi(x)\Phi(-x-\gamma)dx + o(\sigma_\tau^{-1})$$

and

$$\int \Phi(-x)\Phi(x)f(x)dx = f(\mu_\tau) \int \Phi(-x)\Phi(x)dx + o(\sigma_\tau^{-1}).$$

**Proof.**

$$\begin{aligned}
& \left| \sigma_\tau \left[ \int \Phi(x) \Phi(-x - \gamma) f(x) dx - f(\mu_\tau) \int \Phi(x) \Phi(-x - \gamma) dx \right] \right| \\
&= \left| \int \Phi(x) \Phi(-x - \gamma) h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right) dx - h(0) \int \Phi(x) \Phi(-x - \gamma) dx \right| \\
&\leq \int_{x>M} \Phi(x) \Phi(-x - \gamma) h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right) dx + \int_{x<-M} \Phi(x) \Phi(-x - \gamma) h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right) dx \\
&\quad + h(0) \int_{x>M} \Phi(x) \Phi(-x - \gamma) dx + h(0) \int_{x<-M} \Phi(x) \Phi(-x - \gamma) dx \\
&\quad + \int_{|x|\leq M} \Phi(x) \Phi(-x - \gamma) \left| h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right) - h(0) \right| dx \\
&\leq \Phi(-M - \gamma) + \Phi(-M) + h(0) \int_{x>M} \Phi(x) \Phi(-x - \gamma) dx \\
&\quad + h(0) \int_{x<-M} \Phi(x) \Phi(-x - \gamma) dx + \int_{|x|\leq M} \Phi(x) \Phi(-x - \gamma) \left| h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right) - h(0) \right| dx.
\end{aligned}$$

For given  $\gamma$ ,  $\Phi(-M - \gamma)$  and  $\Phi(-M)$  can be arbitrary small for sufficient large  $M$ . Furthermore  $\int \Phi(x) \Phi(-x - \gamma)$  is integrable, and so  $\int_{x<-M} \Phi(x) \Phi(-x - \gamma) dx$  and  $\int_{x>M} \Phi(x) \Phi(-x - \gamma) dx$  can also be arbitrary small for sufficient large  $M$ . For given  $M$ ,  $\int_{|x|\leq M} \Phi(x) \Phi(-x - \gamma) \left| h\left(\frac{x - \mu_\tau}{\sigma_\tau}\right) - h(0) \right| dx$  can also be arbitrary small for sufficient large  $\sigma_\tau$ . So

$$\int \Phi(x) \Phi(-x - \gamma) f(x) dx = f(\mu_\tau) \int \Phi(x) \Phi(-x - \gamma) dx + o(\sigma_\tau^{-1}).$$

Similarly, the other part can be proved.

**Lemma 2.**

$$\int \Phi(-x) \Phi(x + \beta) dx = \beta \Phi\left(\frac{\beta}{\sqrt{2}}\right) + \frac{1}{\sqrt{\pi}} \exp\left\{-\frac{\beta^2}{4}\right\}.$$

**Proof.** By the fact  $d(x\Phi(x) + \phi(x)) = \Phi(x)$  and integration by parts,

$$\begin{aligned}
\int \Phi(-x)\Phi(x + \beta)dx &= \int \phi(x)[(x + \beta)\Phi(x + \beta) + \phi(x + \beta)]dx \\
&= \beta \int \phi(x)\Phi(x + \beta)dx + \int x\phi(x)\Phi(x + \beta)dx + \int \phi(x)\phi(x + \beta)dx \\
&= \beta\Phi\left(\frac{\beta}{\sqrt{2}}\right) + 2 \int \phi(x)\phi(x + \beta)dx \\
&= \beta\Phi\left(\frac{\beta}{\sqrt{2}}\right) + \frac{1}{\sqrt{\pi}} \exp\left\{-\frac{\beta^2}{4}\right\}.
\end{aligned}$$

**Lemma 3.** Suppose  $\sigma_\tau = a\sqrt{n}(a > 0)$  and then

$$\frac{1}{n^{1/4}} \begin{pmatrix} \sum_{i=1}^n [I_{\{d_{i1}=1, d_{i2}=0\}} - EI_{\{d_{i1}=1, d_{i2}=0\}}] \\ \sum_{i=1}^n [I_{\{d_{i1}=0, d_{i2}=1\}} - EI_{\{d_{i1}=0, d_{i2}=1\}}] \end{pmatrix} \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma = \frac{h(0)}{a} \begin{pmatrix} \int \Phi(x)\Phi(-x - \gamma)dx & 0 \\ 0 & \int \Phi(x)\Phi(-x)dx \end{pmatrix}.$$

**Proof:** For  $c_1, c_2 \in R$ , let

$$U_{i\ n} = c_1 [I_{\{d_{i1}=1, d_{i2}=0\}} - EI_{\{d_{i1}=1, d_{i2}=0\}}] + c_2 [I_{\{d_{i1}=0, d_{i2}=1\}} - EI_{\{d_{i1}=0, d_{i2}=1\}}]$$

and then

$$E(U_{i\ n}) = 0, \quad \sqrt{n}E(U_{i\ n}^2) = \frac{h(0)}{a} \left[ c_1^2 \int \Phi(x)\Phi(-x - \gamma)dx + c_2^2 \int \Phi(x)\Phi(-x)dx \right] + o(1).$$

By simple computations,

$$\begin{aligned}
E[\exp\{U_{i\ n}t/n^{1/4}\}] &= 1 + \frac{t^2}{2\sqrt{n}}E(U_{i\ n}^2) + E[o(\frac{U_{i\ n}^2}{n^{1/2}})] \\
&= 1 + \frac{t^2}{2\sqrt{n}}E(U_{i\ n}^2) + o(n^{-1}) \\
&= 1 + \frac{h(0) [c_1^2 \int \Phi(x)\Phi(-x - \gamma)dx + c_2^2 \int \Phi(x)\Phi(-x)dx] t^2}{2an} + o(n^{-1}).
\end{aligned}$$

The moment generating function of  $\sum_{i=1}^n U_{in}/n^{1/4}$  is

$$\begin{aligned}
\phi_n(t) &= E[\exp\{\sum_{i=1}^n U_{in}t/n^{1/4}\}] \\
&= [E(\exp\{U_{in}t/n^{1/4}\})]^n \\
&= \left\{ 1 + \frac{h(0) [c_1^2 \int \Phi(x)\Phi(-x-\gamma)dx + c_2^2 \int \Phi(x)\Phi(-x)dx] t^2}{2an} + o(n^{-1}) \right\}^n \\
&\longrightarrow \exp\left\{ \frac{ah(0) [c_1^2 \int \Phi(x)\Phi(-x-\gamma)dx + c_2^2 \int \Phi(x)\Phi(-x)dx] t^2}{2a} \right\}
\end{aligned}$$

which implies the Lemma holds.

**Lemma 4.** Suppose  $\sigma_\tau = a\sqrt{n}$  ( $a > 0$ ) and the first derivative of  $h(x)$  is continuous, and then

$$n^{1/4} \begin{pmatrix} \frac{\sum_{i=1}^n I_{\{d_{i1}=1, d_{i2}=0\}}}{\sqrt{n}} - \frac{h(0)}{a} \int \Phi(x)\Phi(-x-\gamma)dx \\ \frac{\sum_{i=1}^n I_{\{d_{i1}=0, d_{i2}=1\}}}{\sqrt{n}} - \frac{h(0)}{a} \int \Phi(x)\Phi(-x)dx \end{pmatrix} \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma = \frac{h(0)}{a} \begin{pmatrix} \int \Phi(x)\Phi(-x-\gamma)dx & 0 \\ 0 & \int \Phi(x)\Phi(-x)dx \end{pmatrix}.$$

**Proof:** Since the first derivative of  $h(x)$  is continuous and  $\sigma_\tau = a\sqrt{n}$ , we have

$$\begin{aligned}
\sqrt{n} \times EI_{\{d_{11}=1, d_{12}=0\}} &= \sqrt{n} \times \int \Phi(x)\Phi(-x-\gamma)f(x)dx \\
&= \sqrt{n} \times \int \Phi(x)\Phi(-x-\gamma) \frac{1}{\sigma_\tau} h\left(\frac{x-\mu_\tau}{\sigma_\tau}\right) dx \\
&= \sqrt{n} \times \int \Phi(x)\Phi(-x-\gamma) \frac{1}{a\sqrt{n}} h\left(\frac{x-\mu_\tau}{\sigma_\tau}\right) dx \\
&= \frac{1}{a} \int \Phi(x)\Phi(-x-\gamma) h\left(\frac{x-\mu_\tau}{\sigma_\tau}\right) dx \\
&= \frac{h(0)}{a} \int \Phi(x)\Phi(-x-\gamma) dx + O(\sigma_\tau^{-1}) \\
&= \frac{h(0)}{a} \int \Phi(x)\Phi(-x-\gamma) dx + O(n^{-1/2}).
\end{aligned}$$



Similarly, we can obtain

$$\begin{aligned}
\sqrt{n} \times EI_{\{d_{11}=0, d_{12}=1\}} &= \frac{h(0)}{a} \int \Phi(x) \Phi(-x) dx + O(n^{-1/2}). \\
n^{1/4} &\left( \frac{\sum_{i=1}^n I_{\{d_{i1}=1, d_{i2}=0\}}}{\sqrt{n}} - \frac{h(0)}{a} \int \Phi(x) \Phi(-x - \gamma) dx \right) \\
&\left( \frac{\sum_{i=1}^n I_{\{d_{i1}=0, d_{i2}=1\}}}{\sqrt{n}} - \frac{h(0)}{a} \int \Phi(x) \Phi(-x) dx \right) \\
&= n^{1/4} \left( \frac{\sum_{i=1}^n [I_{\{d_{i1}=1, d_{i2}=0\}} - EI_{\{d_{i1}=1, d_{i2}=0\}}]}{\sqrt{n}} + \sqrt{n} EI_{\{d_{11}=1, d_{12}=0\}} - \frac{h(0)}{a} \int \Phi(x) \Phi(-x - \gamma) dx \right) \\
&\left( \frac{\sum_{i=1}^n [I_{\{d_{i1}=0, d_{i2}=1\}} - EI_{\{d_{i1}=0, d_{i2}=1\}}]}{\sqrt{n}} + \sqrt{n} EI_{\{d_{11}=0, d_{12}=1\}} - \frac{h(0)}{a} \int \Phi(x) \Phi(-x) dx \right) \\
&= n^{1/4} \left( \frac{\sum_{i=1}^n [I_{\{d_{i1}=1, d_{i2}=0\}} - EI_{\{d_{i1}=1, d_{i2}=0\}}]}{\sqrt{n}} \right) + n^{1/4} \left( \sqrt{n} EI_{\{d_{11}=1, d_{12}=0\}} - \frac{h(0)}{a} \int \Phi(x) \Phi(-x - \gamma) dx \right) \\
&\left( \frac{\sum_{i=1}^n [I_{\{d_{i1}=0, d_{i2}=1\}} - EI_{\{d_{i1}=0, d_{i2}=1\}}]}{\sqrt{n}} \right) + n^{1/4} \left( \sqrt{n} EI_{\{d_{11}=0, d_{12}=1\}} - \frac{h(0)}{a} \int \Phi(x) \Phi(-x) dx \right) \\
&= n^{-1/4} \left( \frac{\sum_{i=1}^n [I_{\{d_{i1}=1, d_{i2}=0\}} - EI_{\{d_{i1}=1, d_{i2}=0\}}]}{\sqrt{n}} \right) \\
&\left( \frac{\sum_{i=1}^n [I_{\{d_{i1}=0, d_{i2}=1\}} - EI_{\{d_{i1}=0, d_{i2}=1\}}]}{\sqrt{n}} \right) + o(1)
\end{aligned}$$

which implies the Lemma holds by Lemma 3.

**Proof Theorem 1.** (i) follows immediately from the law of large numbers, Proposition 1 and continuity of  $G(x)$ .

To prove (ii), it follows from the delta method and Lemma 4 above that

$$\begin{aligned}
&n^{1/4} (W - G(\gamma)) \\
&= n^{1/4} \left( \frac{\sum_{i=1}^n I_{\{d_{i1}=1, d_{i2}=0\}} / \sqrt{n}}{\sum_{i=1}^n I_{\{d_{i1}=0, d_{i2}=1\}} / \sqrt{n}} - \frac{\frac{h(0)}{a} \int \Phi(x) \Phi(-x - \gamma) dx}{\frac{h(0)}{a} \int \Phi(x) \Phi(-x) dx} \right) \\
&\xrightarrow{d} N(0, \sigma^{*2})
\end{aligned}$$

where

$$\sigma^{*2} = \frac{a \int \Phi(x) \Phi(-x - \gamma) dx}{h(0) [\int \Phi(x) \Phi(-x) dx]^2} + \frac{a [\int \Phi(x) \Phi(-x - \gamma) dx]^2}{h(0) [\int \Phi(x) \Phi(-x) dx]^3}$$

and then

$$n^{1/4} (\hat{\gamma} - \gamma) = n^{1/4} (G^{-1}(W) - G^{-1}(G(\gamma))) \xrightarrow{d} N(0, \frac{\sigma^{*2}}{[G'(\gamma)]^2}).$$

So

$$\sqrt{\sum_{i=1}^n I_{\{d_{i1}=0, d_{i2}=1\}}} (\hat{\gamma} - \gamma) \xrightarrow{d} N(0, \sigma^2)$$

by

$$\frac{\sum_{i=1}^n I_{\{d_{i1}=0, d_{i2}=1\}}}{\sqrt{n}} \xrightarrow{p} \frac{h(0) \int \Phi(x) \Phi(-x) dx}{a}.$$

**Lemma 5.** Let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \mathbf{x}_{k+1} \in R^k$  satisfy: (a)  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  are linearly independent; (b)  $\mathbf{x}_{k+1} = -c_1 \mathbf{x}_1 - c_2 \mathbf{x}_2 - \dots - c_k \mathbf{x}_k$  where  $c_1, \dots, c_k$  are non-negative real number, and  $r_1, \dots, r_k, r_{k+1}$  be positive real number, then the equation

$$\left\{ \begin{array}{l} G(\mathbf{x}'_1 \boldsymbol{\beta} + \alpha) - r_1 G(-\mathbf{x}'_1 \boldsymbol{\beta}) = 0 \\ G(\mathbf{x}'_2 \boldsymbol{\beta} + \alpha) - r_2 G(-\mathbf{x}'_2 \boldsymbol{\beta}) = 0 \\ \dots\dots\dots \\ G(\mathbf{x}'_k \boldsymbol{\beta} + \alpha) - r_k G(-\mathbf{x}'_k \boldsymbol{\beta}) = 0 \\ G(\mathbf{x}'_{k+1} \boldsymbol{\beta} + \alpha) - r_{k+1} G(-\mathbf{x}'_{k+1} \boldsymbol{\beta}) = 0 \end{array} \right. \quad (15)$$

has a unique solution  $\boldsymbol{\beta}$  and  $\alpha$ .

**Proof:** For fixed  $\alpha$ , let

$$u_\alpha(z) = \frac{G(z + \alpha)}{G(-z)}$$

and

$$\begin{aligned} \frac{du_\alpha(z)}{dz} &= \frac{G'(z + \alpha)G(-z) + G(z + \alpha)G'(-z)}{G^2(-z)} \\ &= -\sqrt{\pi} \frac{\Phi(-(z + \alpha)/\sqrt{2})G(-z) + G(z + \alpha)\Phi(z/\sqrt{2})}{G^2(-z)} \\ &< 0. \end{aligned}$$

So  $u_\alpha(z)$  is decreasing in  $z$  and  $\lim_{z \rightarrow -\infty} u_\alpha(z) = \infty$  and  $\lim_{z \rightarrow \infty} u_\alpha(z) = 0$ . Thus for fixed  $\alpha$ , the equation

$$\begin{cases} G(\mathbf{x}'_1 \boldsymbol{\beta} + \alpha) - r_1 G(-\mathbf{x}'_1 \boldsymbol{\beta}) = 0 \\ G(\mathbf{x}'_2 \boldsymbol{\beta} + \alpha) - r_2 G(-\mathbf{x}'_2 \boldsymbol{\beta}) = 0 \\ \dots\dots\dots \\ G(\mathbf{x}'_k \boldsymbol{\beta} + \alpha) - r_k G(-\mathbf{x}'_k \boldsymbol{\beta}) = 0 \end{cases} \quad (16)$$

has a unique solution when  $\mathbf{x}_1, \dots, \mathbf{x}_k$  are linearly independent.

Let  $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1(\alpha), \dots, \boldsymbol{\beta}_k(\alpha))'$  the solution of (16), and then

$$\frac{d\boldsymbol{\beta}^*}{d\alpha} = -X'^{-1} \delta$$

where

$$\delta = (\delta_1, \dots, \delta_k)', \quad \delta_i = \frac{\Phi(-(\mathbf{x}'_i \boldsymbol{\beta}^* + \alpha)/\sqrt{2})}{\Phi(-(\mathbf{x}'_i \boldsymbol{\beta}^* + \alpha)/\sqrt{2}) + r_i \Phi(\mathbf{x}'_i \boldsymbol{\beta}^*/\sqrt{2})}$$

and

$$X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k).$$

Define

$$t(\alpha) = G(\mathbf{x}'_{k+1} \boldsymbol{\beta}^* + \alpha) - r_{k+1} G(-\mathbf{x}'_{k+1} \boldsymbol{\beta}^*),$$

and then

$$\begin{aligned} \frac{dt(\alpha)}{d\alpha} &= -\sqrt{\pi} \left\{ \left[ \Phi\left(-\frac{\mathbf{x}'_{k+1} \boldsymbol{\beta}^* + \alpha}{\sqrt{2}}\right) + r_{k+1} \Phi\left(\frac{\mathbf{x}'_{k+1} \boldsymbol{\beta}^*}{\sqrt{2}}\right) \right] \mathbf{x}'_{k+1} \frac{d\boldsymbol{\beta}^*}{d\alpha} + \Phi\left(-\frac{\mathbf{x}'_{k+1} \boldsymbol{\beta}^* + \alpha}{\sqrt{2}}\right) \right\} \\ &= -\sqrt{\pi} \left\{ \left[ \Phi\left(-\frac{\mathbf{x}'_{k+1} \boldsymbol{\beta}^* + \alpha}{\sqrt{2}}\right) + r_{k+1} \Phi\left(\frac{\mathbf{x}'_{k+1} \boldsymbol{\beta}^*}{\sqrt{2}}\right) \right] \left( \sum_{j=1}^k c_j \delta_j \right) + \Phi\left(-\frac{\mathbf{x}'_{k+1} \boldsymbol{\beta}^* + \alpha}{\sqrt{2}}\right) \right\} \\ &< 0, \end{aligned}$$

which implies  $t(\alpha) = 0$  has an unique solution and the lemma is proved.

**Proof of Theorem 2.** By Lemma 5 given in the above, it can be proved with  $r_i = p_i/(1 - p_i)$  and  $\mathbf{x}_i = \mathbf{x}_{i2} - \mathbf{x}_{i1}$ .

**Proof of Corollary.** Without loss of generality, suppose that  $\mathbf{x}_{12} - \mathbf{x}_{11}, \dots, \mathbf{x}_{k2} - \mathbf{x}_{k1}$  are linearly independent and

$$\mathbf{x}_{k+1\ 2} - \mathbf{x}_{k+1\ 1} = a_1(\mathbf{x}_{12} - \mathbf{x}_{11}) + \dots + a_k(\mathbf{x}_{k2} - \mathbf{x}_{k1})$$

where  $a_1, \dots, a_k$  is a non-positive real number. Then the determinant

$$\begin{vmatrix} \mathbf{x}'_{12} - \mathbf{x}'_{11} & 1 \\ \mathbf{x}'_{22} - \mathbf{x}'_{21} & 1 \\ \vdots & \vdots \\ \mathbf{x}'_{k2} - \mathbf{x}'_{k1} & 1 \\ \mathbf{x}'_{k+1\ 2} - \mathbf{x}'_{k+1\ 1} & 1 \end{vmatrix} \quad (17)$$

is equal to

$$\begin{aligned} & \begin{vmatrix} \mathbf{x}'_{12} - \mathbf{x}'_{11} \\ \mathbf{x}'_{22} - \mathbf{x}'_{21} \\ \vdots \\ \mathbf{x}'_{k2} - \mathbf{x}'_{k1} \end{vmatrix} \left[ 1 - (\mathbf{x}_{k+1\ 2} - \mathbf{x}_{k+1\ 1})' \begin{pmatrix} \mathbf{x}'_{12} - \mathbf{x}'_{11} \\ \mathbf{x}'_{22} - \mathbf{x}'_{21} \\ \vdots \\ \mathbf{x}'_{k2} - \mathbf{x}'_{k1} \end{pmatrix}^{-1} \mathbf{1}_k \right] \\ &= |\mathbf{x}_{12} - \mathbf{x}_{11}, \mathbf{x}_{22} - \mathbf{x}_{21}, \dots, \mathbf{x}_{k2} - \mathbf{x}_{k1}| \left[ 1 - \sum_{i=1}^k a_i \right] \neq 0 \end{aligned}$$

by the assumption. This implies that the rank of (17) is  $k + 1$ .

Since the rank of  $(\mathbf{1}_m, \mathbf{X}^{*'})$  is equal to that of  $(\mathbf{X}^{*'}, \mathbf{1}_m)$ , which is a  $m \times (k + 1)$  matrix, and (17) is a matrix obtained by the first  $k + 1$  rows of  $(\mathbf{X}^{*'}, \mathbf{1}_m)$ , thus the rank of  $(\mathbf{1}_m, \mathbf{X}^{*'})$  is  $k + 1$ .

## References

- Arellano, M. (2003). Discrete choices with panel data. *Investigaciones Economicas*, 27, 423-458.
- Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, 77, 489-536.
- Arellano, M. and Honore, B. (2001). Panel data Models: some recent developments. *Handbook of Econometrics*, Vol. V, ed. by J. Heckman and E. Leamer. Amsterdam: North Holland.
- Bartolucci, F. and Farcomeni, A. (2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104, 816-833.

- Bartolucci, F. and Nigro, V.(2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a  $\sqrt{n}$  consistent conditional estimator. *Econometrica*, 78, 719-733.
- Chamberlain, G.(1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, 225-238.
- Chamberlain, G.(1985). Heterogeneity, omitted variables bias, and duration dependence. *Longitudinal Analysis of Labor Market Data*, edited by Heckman, J. and Singer, B. Cambridge University Press.
- Heckman, J.(1978). Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales de l'INSEE* 30/31, 227-269.
- Heckman, J.(1980). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. *Structural Analysis of Discrete Data with Econometric Applications*, ed by C. F. Manski and D. McFadden, p179-195. Cambridge, MA: MIT Press.
- Heckman, J.(1981). Heterogeneity and state dependence. *Studies in Labor Markets*, ed by S. Rosen, p91-140. University of Chicago Press.
- Hisao, C.(2003). *Analysis of Panel Data*(Second Ed.). New York: Cambridge University Press.
- Honore, B. and Kyriazidou, E.(2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68, 611-629.
- Horowitz, J. L.(1992). A smoothed maximum score estimator for binary response model. *Econometrica*, 60, 505-531.
- Lancaster, T.(2000). The incidental parameter problem since 1984. *Journal of Econometrics*, 95, 391-413.

- Lancaster, T.(2002). Orthogonal parameters and panel data. *Review of Economic Studies*, 647-666.
- Manski, C. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 55, 357-362.
- McCullagh, P. and Nelder, J. A.(1989). *Generalized Linear Models*. London: Chapman & Hall.
- Neyman, J. and Scott, E. S.(1948). Consistent estimation from partially consistent observations. *Econometrica*, 16, 1-32.